Hydrologcial modelling within GIS: an integrated approach

Citation for published version:
Stuart, N & Stocks, C 1993, 'Hydrologcial modelling within GIS: an integrated approach' Paper presented at HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources, Vienna, Austria, 1/04/93 - 1/04/93, pp. 319-329., No DOI for this publication

Digital Object Identifier (DOI):
No DOI for this publication

Hydrological modelling within GIS: an integrated approach

NEIL STUART & CHRISTOPHER STOCKS

Department of Geography, University of Edinburgh, Drummond Street, Edinburgh, Scotland EH8 9XP, UK

Abstract Incorporating tools for physically based hydrological modelling within GIS offers mutual benefits. The presently favoured semi-distributed modelling approach seems particularly appropriate for closer linkage with GIS. Whilst specific hydrological models are best linked to GIS loosely through an interface, integrating a set of generic modelling tools within a GIS can create an advantageous environment for model development. To illustrate this, a set of tools are embedded within one commercial GIS and used to implement a semi-distributed hydrological model. Results show that this integrated approach has the advantage of producing a single environment in which users can conduct all their modelling work. This suggests that as GIS become more flexible, supporting a wider range of data models and more sophisticated customization languages, they may well become the preferred environment for hydrological modelling.

INTRODUCTION

Progress in hydrological simulation modelling and the development of geographical information systems (GIS) have largely occurred on parallel, but clearly separated tracks. This separation is surprising when one considers the common ground that the two research areas share and the scope for developments in each area to offer benefits to the other.

In the last ten years GIS have developed to the stage where large databases containing a diversity of geographic information in point, line, polygonal and cellular formats can be integrated, selectively manipulated and related to derive continuous estimates of many environmental parameters over a range of scales. Yet many GIS are still being used as advanced digital cartographic systems, oriented towards the maintenance of digital geographic data and relatively weak in respect of the higher analysis and modelling abilities that are required to make them effective decision support systems (Kehris, 1990; Densham, 1991).

Over a similar time period, researchers attempting to model hydrological processes at the scale of the drainage basin have found difficulty in synthesizing and aggregating data from a limited number of locations and in finding data of an appropriate quality and resolution for "fully distributed" models, such as the Institute of Hydrology Distributed Model (IHDM) or the Syteme Hydrologique European (SHE). It seems that the fine spatial resolution at which these models operate and their need for many parameter values which are difficult to measure in the field has limited their usefulness for practical purposes (Anderson & Rogers, 1987; Bathurst, 1988).

Partly because of these difficulties, the "semi-distributed" approach to hydrological modelling has gained ground as an attractive means of retaining the benefits of physically realistic results, but without these overheads of fully distributed models (Beven, 1989). The approach is more selective in the number of environmental variables employed and more sensitive in choosing an appropriate resolution for spatial partitioning. Some of the more attractive features of the semi-distributed approach include:

(a) adoption of a "minimalist" philosophy towards data selection. The importance of all available data is assessed and only the key variables are used as inputs to a model;
(b) simplicity is emphasized to retain understanding; only a subset of major processes are modelled by simple equations;
(c) the landscape is partitioned into units which reflect the spatial variability in the main driving processes and controls. The partitioning need not be into uniform cells, but into "response units" which have some hydrological significance or uniformity of process;
(d) an exploratory approach to model building is envisaged. Starting from a simple prototype, the effects of adding or removing variables can be explored to understand the sensitivity of the model. Also, the model can be run with different spatial partitionings and the results compared.

A CASE FOR CLOSER LINKAGE OF MODELS WITH GIS

The "semi-distributed" distributed approach has close parallels to methods used by workers developing GIS applications. A case for more closely linking model building to GIS can be made by identifying several convergent approaches in the modelling of data, in the partitioning of space and in exploring spatial data.

The concept of selecting key spatial variables such as the modelling substrate (e.g., elevation) and environmental controls on hydrological process (e.g., vegetation interception or soil porosity) is similar to the GIS concept of building a model of geographical space in which certain themes or layers are of interest.

Considering the spatial data models used in both fields, there is an obvious similarity between the "bottom up" approach of fully distributed models and GIS based on the raster or cellular data model; in both cases the area of study is partitioned at a fixed resolution by a fine mesh. The regular unit leads to the advantage of computing simplicity but the drawback of large volumes of data. On the other hand, the semi-distributed paradigm recognizes the idea of patches of the landscape within which conditions can be considered relatively uniform. This higher level identification of patches is similar to the notion of objects or features in some GIS. The higher level view has the advantage of simplicity in that we naturally classify a landscape into features and of efficiency in that grouping reduces the volume of data to be stored and processed.

A crucial aspect of semi-distributed modelling is the partitioning of a landscape based on a "response unit" which reflects local variations in the environmental conditions. The choice of variables and the appropriate spatial and temporal resolutions forms the structural basis of a model and can fundamentally affect the nature and reliability of results produced (Anderson & Burt, 1985). As Fig. 1 indicates, the concepts of classification and overlay can be used in a GIS to experiment with different layers and different spatial resolutions to derive a meaningful "response unit".

From an analysis of these convergent approaches it seems that a closer integration of GIS and hydrological modelling could be mutually beneficial. From a GIS perspective, the ability to characterize and model the spatial variations in hydrological processes would make GIS a more effective aid for planning and managing the use of land within a drainage basin. From a modelling perspective, GIS allow environmental data from disparate sources to be georeferenced and related; missing or discontinuous data values to be interpolated and many different combinations and resolutions of geographic data to be explored.

Whether or not these benefits lead one to conclude that GIS is only a complementary tool for model building, or in fact the possible future platform for model development, seems to depend on the human and technical obstacles to linking the two.

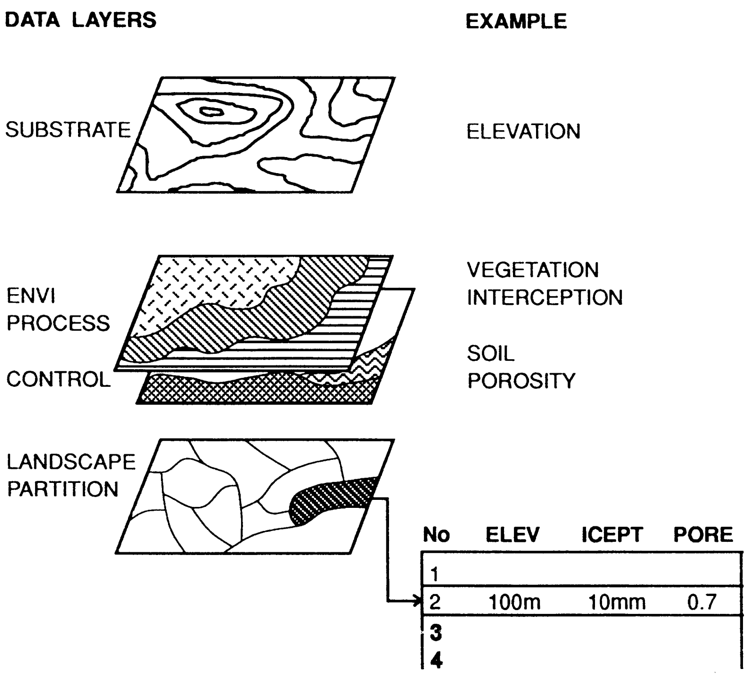

Fig. 1 Using GIS to overlay selected, classified data layers of hydrological significance, different spatial partitions of a catchment can be explored.

Opportunities and constraints on making the link

Until recently, GIS have been restricted in their complexity of computation to the mathematical operators offered by standard DBMS for performing calculations on attribute tables. This syntax limits the implementation of modelling equations to modest calculations of a few lines of code.

However, as GIS become more "open" and users demand greater ability to interact with them, more versatile customization languages are appearing, which allow users access to the underlying data structures and procedures. With these languages gaining more features of full programming languages, such as variables and looping constructs, it becomes possible to embed general purpose modelling tools within a customized GIS.

Some of the reasons for the separate development of GIS and hydrological models may therefore be understood as a previous inability of GIS

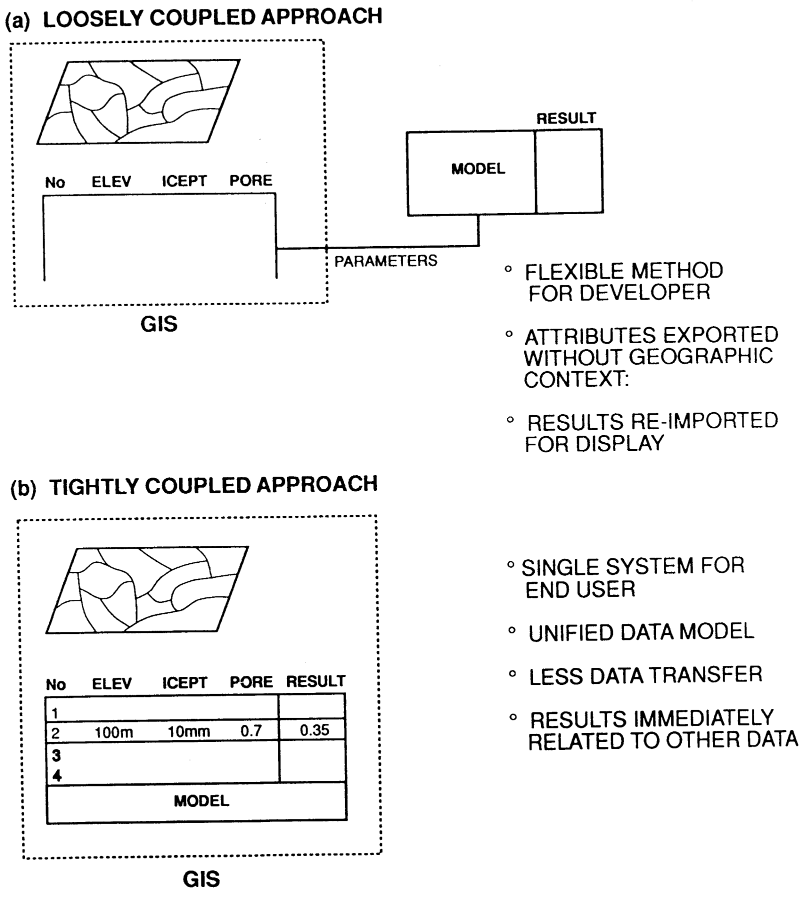

Fig. 2 Two alternative ways of linking a model to a GIS; in (a) the model is linked loosely through an interface; in (b) the model is encoded within the GIS and directly accesses the data structures and procedures of the GIS.

to meet the needs of the model builder for appropriate data models and flexible program development. In bringing modelling and GIS together at this time, two alternative strategies can be identified as shown in Fig. 2.

There are several well established hydrological models, implemented in standard programming languages. These may wish to use GIS for preparing and synthesizing data gathered in different forms and from different areal units and for experimenting with alternative spatial partitions. However, the model will still run externally, since the algorithms often exploit structures of the host programming language. Some recent prototypes are reported by Mallents & Badji, (1991) and Foster (1991). The investment in existing coding implies that a loose coupling to GIS is the best solution, at least in the short term. As Fig. 2a shows, the GIS supplies parameter estimates which are exported to the model; after performing the calculations, results that are locationally referenced may be reimported into the GIS for visualization.

In this case, each system is used for its designed purpose, but considerable engineering is required to bolt the two together; in most cases the link is not seamless as one system must be exited before the other can be run; Effort is needed to maintain the link, as there is no guarantee that the two packages will maintain their internal structures as software is upgraded. Often only the attribute data can be easily exported as ASCII tables, with the geographical context of the data sometimes being lost.

Whilst this method may suit researchers, end users with less technical expertise wish for a single environment and a unified data model as shown schematically in Fig 2b. Here, the user has more time available for developing and refining an appropriate model and wastes less time on the mechanics of data transfer. If a set of general purpose tools for model building can be provided within a system, then new models can be rapidly prototyped. To the user, this customized package seems more like a model development environment than a GIS.

A CASE STUDY ILLUSTRATION

To show how a set of general modelling tools can be embedded within a commercial GIS, using the functionality of an internal customization language, a set of tools have been written using the modelling language of the SPANS GIS. These tools are used to implement elements of the semi distributed TOPMODEL for predicting surface saturation. The test area selected is the Gwy catchment in west Wales. This is a small headwater catchment of some 4 km^2 and with a total stream length of 5.7 km. The catchment lies between 340 and 740 m and is distinctively upland; steep grassy slopes and brown earth soils mean that water is typically transmitted by lateral throughflow, often through an extensive network of natural soil pipes. The annual precipitation is around 2500 mm. The catchment was chosen for reasons of data availability and for continuity with previous studies. As the catchment has been instrumented by the UK Institute of Hydrology since 1968, there is a good supply of hydrological data. The land cover and soils within the catchment have also been mapped in some detail.

Model design

The hydrological algorithms used in this study were adapted from TOPMODEL (Beven & Wood, 1983; Beven 1989). This was selected for implementation because it uses a hydrologically proven, semi-distributed approach; its mathematics are relatively straightforward and well documented and its data demands modest, making it ideal for a modelling environment based on GIS. Also, TOPMODEL has previously been applied to the catchment, so key parameter values that require specialist calculation are available. Finally, as a result of previous work in Edinburgh by Stuart & Hartshorne (1992), digital datasets needed by TOPMODEL existed for the site in formats which were suitable for import into SPANS.

Using the basic TOPMODEL theory that explicitly links topographic form to subsurface water flow and the production of surface runoff, a set of physically based hydrological simulation tools have been implemented within SPANS. A catchment is partitioned on the basis of a topographic index describing hillslope form and further subdivided according to soil and vegetative conditions. The model is driven from precipitation records and the resulting patterns of surface saturation are modified by estimates of the spatial variations in evapo-transpiration, soil transmissivity and areas where pipe flow is known to occur (a factor frequently neglected in previous models).

Since the theoretical basis of TOPMODEL has been clearly reported by its authors (Beven & Kirkby, 1979; Beven & Wood, 1983), the example given below emphasizes how one of the principal equations, for deriving local subsurface flow, may be incorporated within a GIS and run upon a suitably constructed partitioning to visualize the spatial variations in subsurface flow and water deficit across a catchment.

Consider a segment of hill slope, draining through point (i) in a catchment. Assume that the subsurface flow rate (qi) can be related to soil moisture deficit (Si) by the relationship:

$$qi = To \tan(B) \exp(-Si/m) \quad (1)$$

where (tan (B)) is the slope and is taken as an approximation to hydrological gradient at point (i), (To) is a soil conductivity or transmissivity parameter when the profile is just saturated (i.e. $Si = 0$) and (m) is a parameter relating to the rate of change of conductivity in the profile. For the case of a steady input rate r to the slope (precipitation intensity), then at any point the down slope flow must be given by:

$$qi = ar \quad (2)$$

where (*a*) is the area of a topographic partial contributing area upon which precipitation is falling. The combination of equations (1) and (2):

$$Si = -m \ln(ar / To \tan(B) \quad (3)$$

provides the basis for the calculation of distributed local saturation deficit. Implementing this equation within SPANS is a two stage process; firstly distributed parameter values must be provided at the appropriate spatial resolution, using the GIS to derive and overlay multiple maps. Note that a map in the GIS is synonymous with a layer containing spatially distributed parameter values. Secondly the parameter values must be accessed by a modelling routine written in the internal Spans Modelling Language (SML).

A distributed hydrological database was established to provide common combinations of parameter values likely to be required for the modelling equations. Figure 3 summarizes the map data layers held as quadtree maps (.MAP) and the associated attribute tables (.TBB) which are linked to them. An attribute table is a look-up table, used to hold a range of hydrological data sets, such as rainfall records for different time periods.

The spatial partitioning of an area into units upon which a modelling equation can operate is achieved by overlaying simultaneously all maps containing parameter values required by the model. The subsequent map produced is partitioned into thousands of tiny areas; each area contains an individual combination of constituent variables from the overlain layers. Such a map is termed a "unique conditions map". Associated with it is a large attribute table, indicating for each unique area on the map, the class on each of the input maps from which it was derived. This map is used for modelling, rather than visualization, since it enables each unique area to be visited in turn and the value of each of the required parameters at that location to be extracted (or looked up). An overlay of the most important data sets for modelling saturation deficit involved combining eight maps to produce a "master" map and table of unique conditions. The function of this map and table as the common link between all other maps and their attribute tables is shown schematically in Fig. 3.

Hence, to evaluate equation (3) above across the catchment, minimally requires the overlay of the following five maps:

(a) partial contributing areas (part.map) for "a";
(b) local slope tangent (slopelut.map) for "tan(B)";
(c) soil transmissivity (soillut.map) for "To";
(d) rainfall intensity (rainfile.map) for "r";
(e) parameter m (veglut.map) for "m".

Executing the model can be represented simply by the following algorithm, where CALC implies a calculation on a raster map (in a style afte Kwadijk & Van Deursen, 1990):

(a) TIME STEP;
(b) CALC Input.map = part.map x rainfile.map;
(c) CALC Substrate.map = soillut.map x slopelut.map;

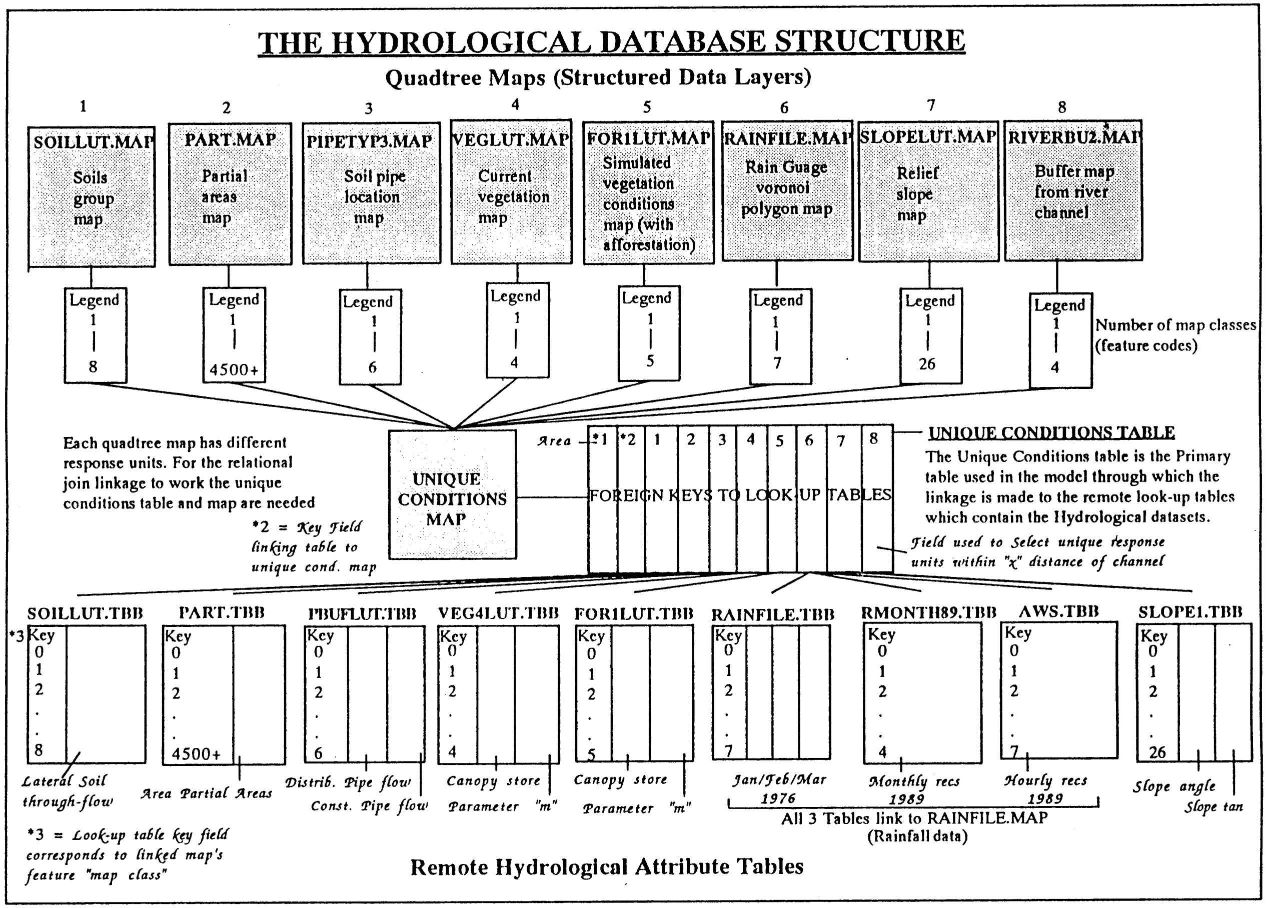

Fig. 3 The structure of the GIS database required for executing a range of hydrological modelling tools; Eight map layers (.MAP), each with one or more attribute tables (.TBB) are linked together through the composite unique conditions map.

(d) CALC Deficit.map = -(veglut.map) x Ln(input.map/substrate.map).

This expresses the idea of visiting each location, looking up values on a number of maps at that point and computing a modelling result. In fact it is actually the numeric values held in associated attribute tables that are being accessed and operated upon and a result is computed once for each unique area, not for each single cell.

Model results

As a result of this work, a toolbox of SML procedures has been produced. The tools presently implemented are those for retrieving and manipulating data and those for performing calculations according to a selected algorithm. For this example, whilst the algorithm is based on TOPMODEL, there is considerable flexibility in the selection and derivation of parameters. For example, facilities are included to derive a value for To at run time, representing either soil matrix flow alone, or including distributed pipe flow.

The results of running a modelling equation with and without the assumption of pipeflow can be assessed qualitatively by comparing the resulting outputs of Figs 4 and 5. In Fig. 4, maximum values for Si, approximating to areas of greatest throughflow are closely related to areas of steepest slope, which in this catchment are quite close to the channel. Figure 5 shows that by recalculating to account for the spatial distribution of soil pipes, many areas that could contribute to rapid throughflow expand and coalesce, which could markedly affect the hydrological response of the catchment.

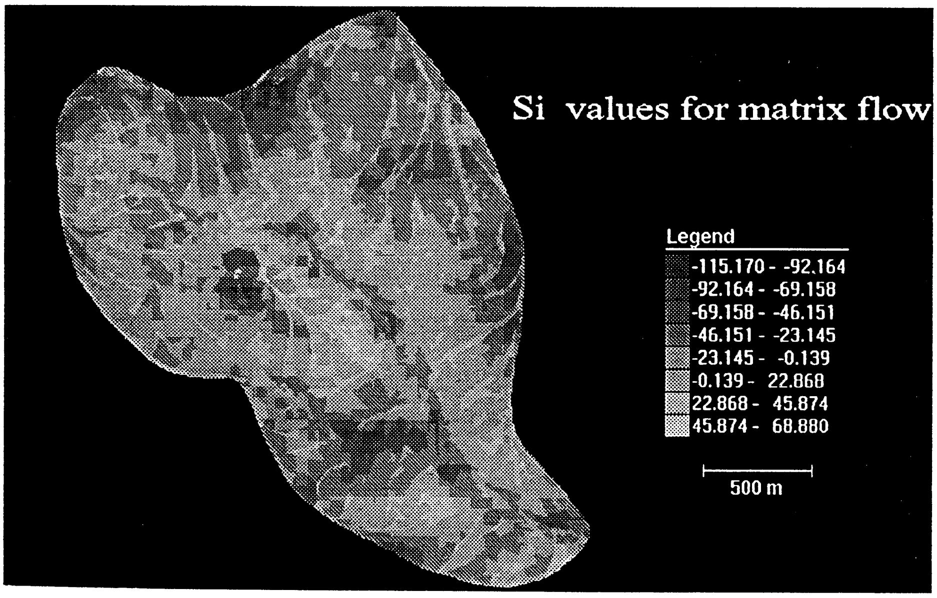

Fig. 4 Spatial distribution of saturation deficits in the Gwy catchment, estimated using TOPMODEL assuming no pipe flow.

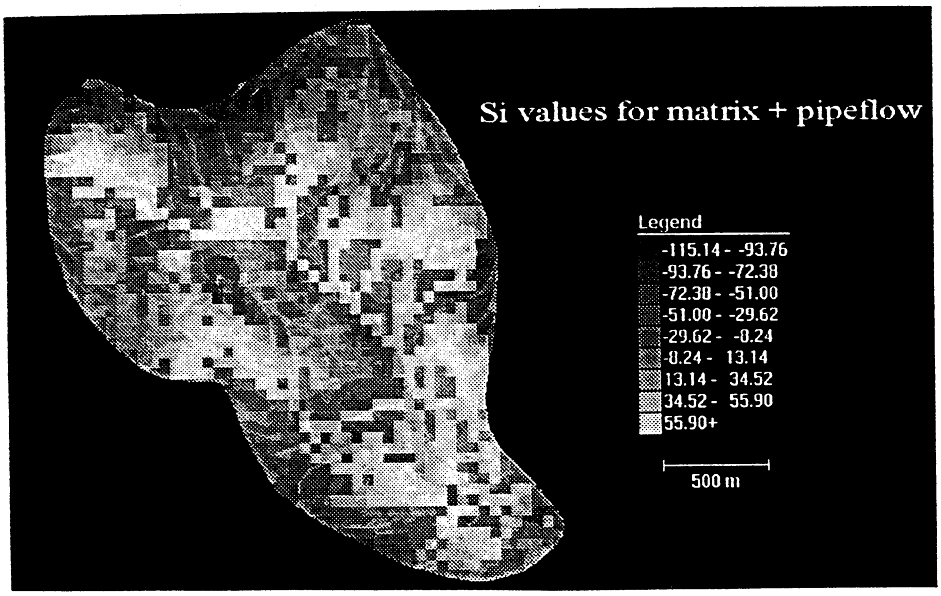

Fig. 5 Spatial distribution of saturation deficits in the Gwy catchment, assuming all soil pipes connected and flowing.

Individual tools in the form of blocks of code can be combined in different ways to produce variant models on a common theme. Other tools allow a user to define at run-time which pipe areas are to be simulated as ephemeral, which as perennial. Actual flow rates may be specified. There is a facility to simulate afforestation of the catchment and to explore how changes of land cover may affect the pattern of moisture deficits.

DISCUSSION AND CONCLUSIONS

The set of tools implemented are still simplistic in a number of respects. Most oversimplification results from the constraints of working within the limitations of the internal modelling language, which is still not as flexible as a high level programming language. Some examples of weaknesses are the lack of a looping construct, a limited number of variables and a limited precision for calculating with real arithmetic. As a result, the outputs produced may be interpreted as indicative, but without reliance on the absolute values being produced.

Whilst further work is required to validate and calibrate these models, this example seeks mainly to demonstrate some of the attractive concepts of integrating modelling tools within GIS. A single environment is created; within which all stages of model building can be undertaken from the initial selection and transformation of data sets, to the final visualization of results in their geographic context.

The concept of unique conditions modelling is similar to idea of the "response unit" in hydrological modelling except that being in a GIS the exact locations of the unit are known; the unique area contains all the parameter values required to evaluate a hydrological model for that location and the shape and size of the unit is also dependant on the layers of data selected. Models based on a few, generalized parameter maps will have few unique conditions, be quick to operate, but produce smoothed results. In comparison, using many parameters involves overlaying many maps and will produce a very complex mosaic that is very slow to run a model upon. This method therefore offers a flexible environment in which to experiment iteratively, seeking to find that optimal balance: a minimal set of maps which produces physically realistic results without the problems of "over parameterization" (Beven, 1989).

The investment in existing code means that many hydrological models will only take advantage of interaction with GIS through a loose form of linkage or interfacing. In the longer term, as GIS support a wider range of data structures and become open systems, bound together by more flexible internal languages, the challenge will be for hydrologists to customize GIS into environments for model development. The prototype set of general modelling tools illustrated here are hopefully a small step in this direction.

REFERENCES

Anderson, M.G. & Burt, T.P. (1985) Modelling Strategies. In: *Hydrological Forecasting* (ed. by M.G. Anderson & T.P. Burt), Wiley, chapter 1.

Anderson, M.G. & Rogers, C.C.M. (1987) Catchment scale distributed hydrological models: a discussion of research directions. *Progress in Physical Geography* **11**(1), 28-52.

Bathurst, J.C. (1988) Physically based distributed modelling of an upland catchment using the Systeme Hydrologique European. *J. Hydrol.* **87**, 79-102.

Beven, K.J. & Kirkby, M.J. (1979) A physically based variable contributing area model of basin hydrology. *Hydrol. Sci. Bulletin* **24**(1) 43-69.

Beven, K.J & Wood, E.F. (1983) Catchment geomorphology and the dynamics of runoff contributing areas. *J. Hydrol.* **65**, 139-158.

Beven, K.J. (1989) Changing ideas in hydrology - the case of physically based models. *J. Hydrol.* **105**, 157-172.

Burrough, P.A. (1989) Matching spatial databases and quantitative models in land resource assessment. *Soil Use and Management* **5**, 3-8.

Densham, P.J. (1991) Spatial decision support systems. In: *Geographical Information Sytems: Principles and Applications* (ed. by D.J. Maguire, M.F. Goodchild & D.W. Rhind) Longman, vol. 1, chapter 24, 403-412.

Foster, G.C. (1991) Integration of a semi-distributed model in a GIS environment. MSc. Dissertation, University of Edinburgh.

Kehris, E. (1990) A geographical modelling environment built around ARC/INFO. *Proc. EGIS 1990*, 556-564.

Kwadijk, J. & Van Deursen, W. (1990) Using the WATERSHED tools for modelling the Rhine catchment. *Proc. EGIS 1990*, 255-263.

Mallants, D. & Badji, M. (1991) Integrating GIS and deterministic hydrological models: a powerful tool for impact assessment. *Proc. EGIS 1990*, 672-679.

Mitchell, B. (1991) *Geography and Resource Analysis*. Longman, Chapter 6.

Stuart, N. & Hartshorne, J. (1992) Hydrological modelling with GIS - approaches and opportunities. *Proc. First European SPANS User Conference*, Amsterdam, 45-57.